Prisoner's dilemma and Nash Equilibrium

Alice and Bob were just caught transferring state secrets (darn those bad generators!). Now, sadly, they face prison time. Separated into 2 rooms, homeland security tries to get them to confess. They are each told (independently) that if they both confess, they will be put in prison for 3 years. If one confesses and the other does not, the confessor will be let free in exchange for testifying against the other, who will receive 4 years in prison. If they both keep quiet they will be let off with a slap on wrist: 1 year each. The outcomes are represented in the following table:

	C	D
C	3 , 3	0 , 4
D	4 , 0	1 , 1

Def 1.1 We say that a collection of players strategies form a **Nash Equilibrium** if no deviation in any single player's strategy results in a higher yield for that player.

Let us look at the prisoner's dilemma with this definition in mind. It has one equlibrium: the case where both players confess. If Alice knows that Bob will confess, then she cannot do anything but confess as well: a deviation from the equilibrium can only result in a fourth year for her. Of course it is possible that both players will choose not to confess from the get go, but this is **not** an equilibrium. If Alice gets wind of the fact that Bob plans to stay quiet, she'll turn him in! And if bob in turn realizes this, he'll choose to confess as well. So while this can occur, in some sense it is an "unstable" outcome.

Let's turn to another example: the lion and the lamb. Two predators are fighting over a piece of meat. If they both chill and play the lamb, they can split it. We'll say they both gain 3 points. If one plays lion while the other plays lamb, he makes out with the meat gaining 4 points while his opponent gets the bones (worth 1 point). If they both play lion they both lose out and receive no points:

	Lamb	*Lion*
Lamb	3 , 3	1 , 4
Lion	4 , 1	0 , 0

This game has 2 equilibrium: (Hawk, Lamb) and (Lamb, Hawk).

Before we move on, let's look at one more example called matching pennies. See if you can find the equilibrium:

Heads *Tails*

Heads	1 , −1	−1 , 1
Tails	−1 , 1	1 , −1

As we see, not all games have a Nash equilibrium. Even when they do, we should recognize that this equilibrium does not necessarily predict what will happen. Consider the game stop/don't stop. I ask each of you in succession whether you'd like to stop, or continue. If anyone ever says stop, I give them $10. However, if every player says continue, then at the end I give **every** player $11. The strategy profile where everyone says continue is an equilibrium. However, if we consider what would happen in real life, it is quite likely someone would give up the dollar and stay stop. Especially if there were 100 students in line to be asked next. So what gives? Why does the theory fail here? Well, we simply haven't evaluated the payoffs correctly. In terms of concrete dollars it is correct. But the value of money in the hand is greater than the value of money to come. Most people would pay $1 to assure $10. How much this assurance is worth is hard to evaluate. In fact, it probably varies from one person to the next. But what if the game were played with $2 to the one that stopped and $1000 to everyone if all say continue? Then probably very few players would stop. Finding a good way to model real life situations can be difficult. But by simplifying problems we can learn a lot about the economics of human (and even non-human) interaction, and this is what game theory attempts to do.

1 Mixed Strategies

Let's consider one more game called *BoS* (originally Bach or Stravinsky). Alice and Bob are celebrating their freedom with a movie. Bob wants to go see Batman, but Alice really prefers Something About Love (1988). They do agree on one thing: they want to see the movie together. They can't hold hands if they're in different theaters. Here are their payoffs:

	B	*S*
B	2 , 1	0 , 0
S	0 , 0	1 , 2

We know already that this game has two equilibrium: (*B*,*B*) and (*S*,*S*). But in fact there is another less obvious equilibrium. The two more obvious equilibrium are called "pure-strategy" equilibrium, but it is also possible for the players to choose a "mixed-strategy" in which (you guessed it) they will play *B* with some probability and *S* with the remaining probability. The goal is to find the appropriate probabilities such that neither player will choose to strange their strategy. More specifically, let's define $p_1(B)$ and $p_2(B)$, to be the probabilities that player 1 chooses B and that player 2 chooses B respectively. We want to find values of $p_1(B)$ such that player 2 becomes indifferent between the choice of B and S (we will then do the same for player 1):

$$2 \cdot p_1(B) + 0 \cdot p_1(S) = 0 \cdot p_1(B) + 1 \cdot p_1(S)$$

$$\Rightarrow 2 \cdot p_1(B) = 1 \cdot (1 - p_1(B))$$

$$\Rightarrow p_1(B) = \frac{1}{3}$$

A similar equation to calculate $p_2(B)$ yields $\frac{2}{3}$.

2 Dominance

Sometimes players can rule out certain strategies pretty easily. We call such strategies "dominated" strategies, because in any situation it does not pay to play such a strategy. If we remove all of these strategies, however, we are left with a different game in which some **other** strategies may in fact be removable. By doing this over and over again, we can rule out a lot of strategies. Consider the following game:

	L	R
U	4 , 0	1 , 1
M	0 , 0	3 , 1
D	1 , 5	1 , 0

We can see that there is no reason for player 1 to play D, because U is always at least as good as D. But once player 2 realizes this, he may as well always play R. Now player 1 concludes that if player 2 plays R, he may play M. So (M,R) is the strategy profile that survives iterated deletion of weakly dominant strategies. We call this a *rationalizable* strategy.